Synthetic Surface Rainfall Radar Retrieval Using Conditional GAN Algorithm

Shujia Zhou

Xiaowen Li

Dan Duffy

NASA GSFC NCCS

Shujia.zhou-1@nasa.gov

Goal

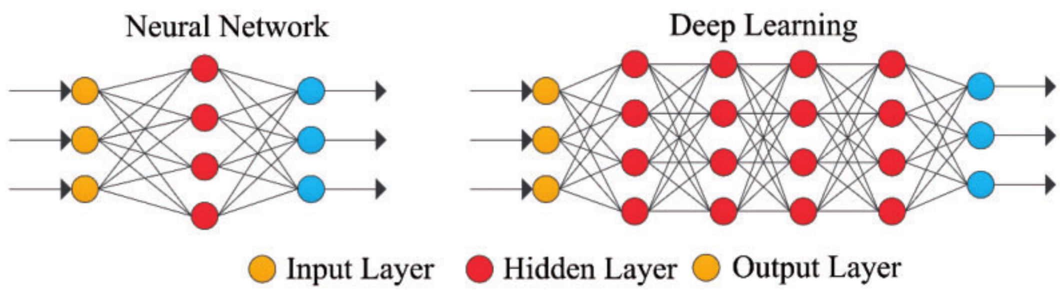
- Use Machine Learning/Deep Learning Algorithms to develop a model of retrieving instantaneous surface rainfall

Approach

- Experiment with cloud-resolving model simulated radar reflectivity and surface rainfall rate.
  - The advantage of using synthetic data is that we have a well-behaved dataset, free of instrument calibration errors and sampling/matching errors.
- Use the state-of-the-art deep learning algorithm, conditional Generative Adversarial Nets (cGAN).

Machine Learning/Deep Learning

Deep Learning	Applications
Multi layer neural network	General
Long term short memory (LSTM)	Dynamic/time dependence
Convolutional Neural Network (CNN)	Spatial dependence
Auto encoder	Dimensionality reduction
Generative adversarial network (GAN)	Variable-to-variable translation. Deep fake



Turning Parameters in deep learning

- Number of layers
- Number of nodes per layers
- Number of EPOCH
- Error metrics
- Loss function

Illustration of Generative Adversarial Network (GAN) Applications

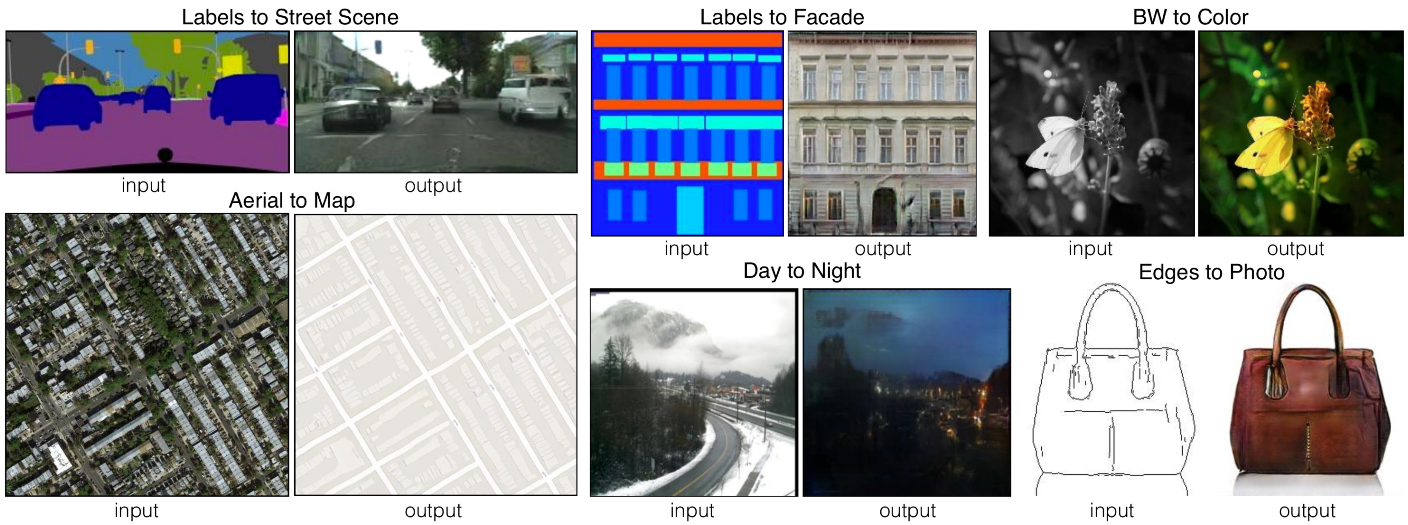
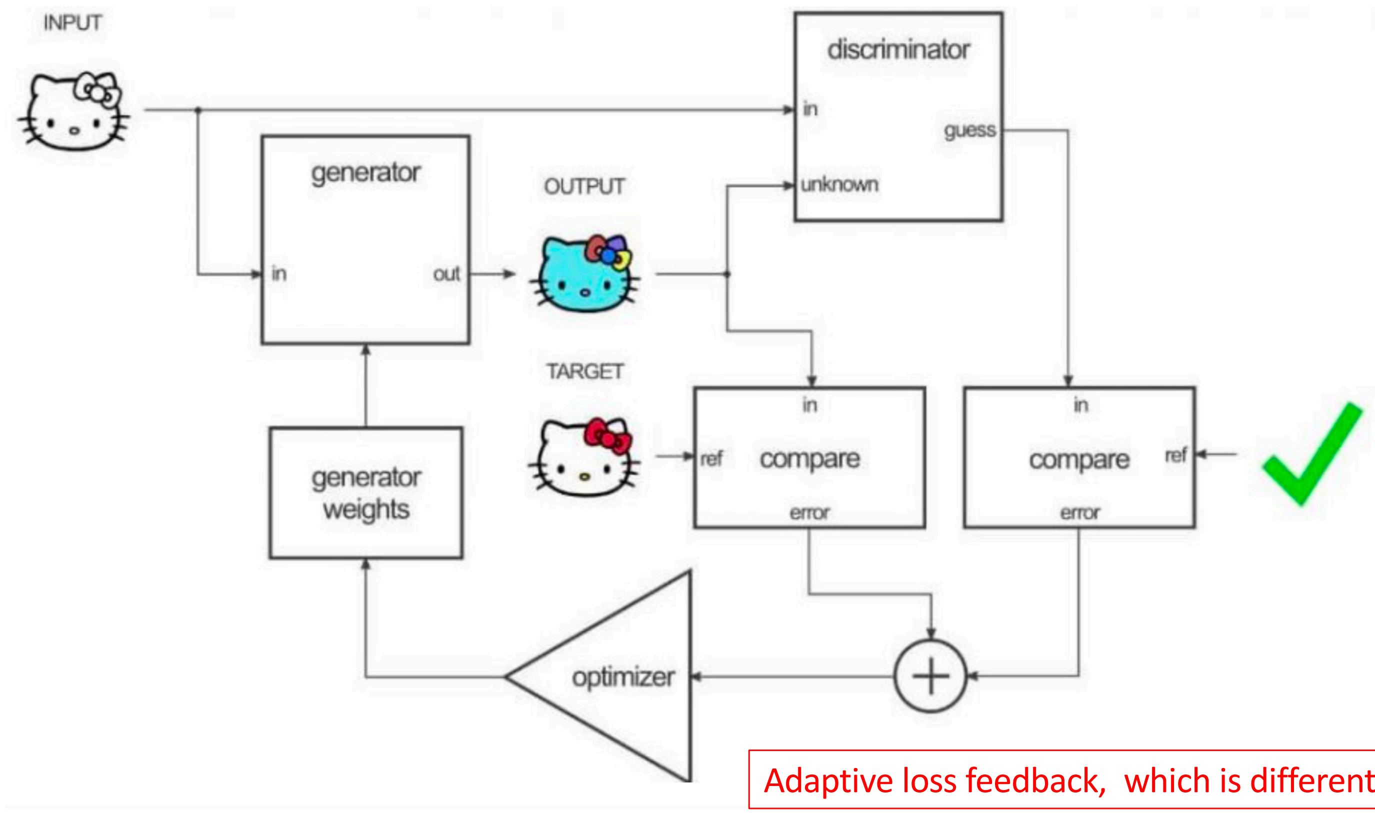


Figure 1: Many problems in image processing, graphics, and vision involve translating an input image into a corresponding output image. These problems are often treated with application-specific algorithms, even though the setting is always the same: map pixels to pixels. Conditional adversarial nets are a general-purpose solution that appears to work well on a wide variety of these problems. Here we show results of the method on several. In each case we use the same architecture and objective, and simply train on different data.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *arXiv preprint*, 2017

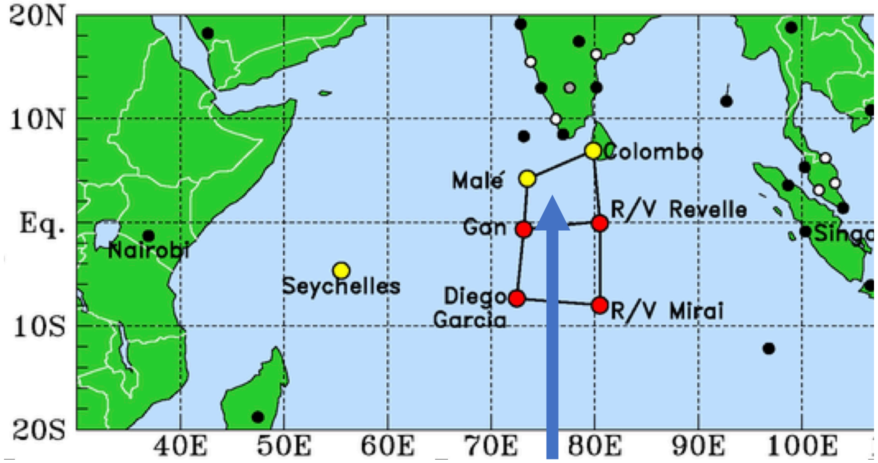
Conditional GAN Architecture



Adaptive loss feedback, which is different from CNN

Goddard Cumulus Ensemble (GCE) Model Simulation

30-day simulation over the tropical Indian Ocean during November 2011 Madden-Julian Oscillation event. Hourly surface rainfall output as shown in the left panels.



Model Simulation Domain (256 x 256 grid points)

Simulation Data

- The radar and rainfall data are from GCE simulations. There are no instrument-related error.
- Training data: 480 images in jpeg (Nov. 1 to Nov. 20)
  - Shuffled during the training---no time dependency
- Test data: 240 images in jpeg (Nov. 21-30)
- Each image consists of input (radar reflectivity) and target (rainfall rate) images in 256x256 pixels

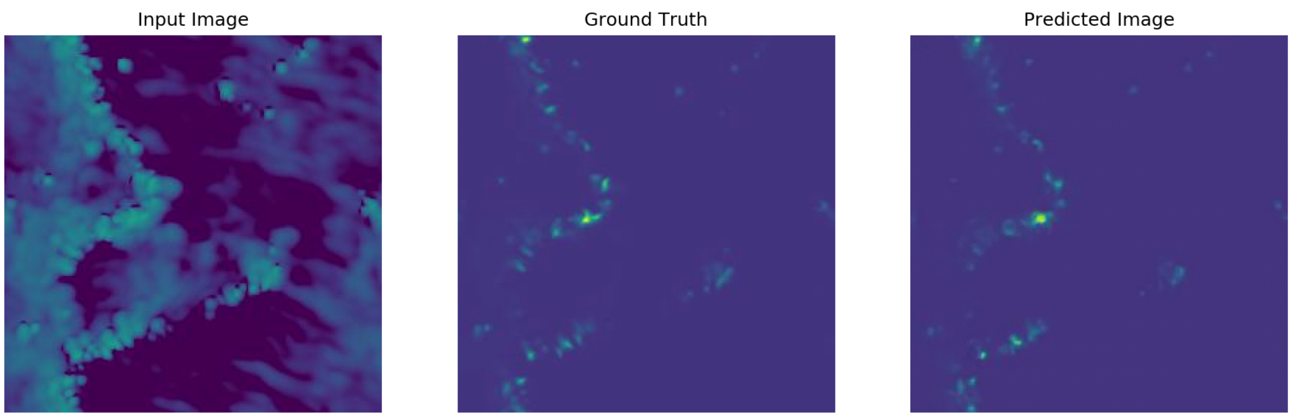
Simulation Environment

- NASA NCCS GPU platforms
  - NVIDIA V100 with 8 GPUs or 4 GPUs
- TensorFlow 1.x, 2.0 were used
  - Pix2pix code was used and modified for this simulation
    - Pix2pix uses CNN
  - Code development and testing run were in Mac Notebook with CPU
  - Production run were in GPU V100
    - TensorFlow supports multi-GPU run but the performance was not improved much over that of a single GPU. It is very likely due to the inefficient interconnection network among GPUs

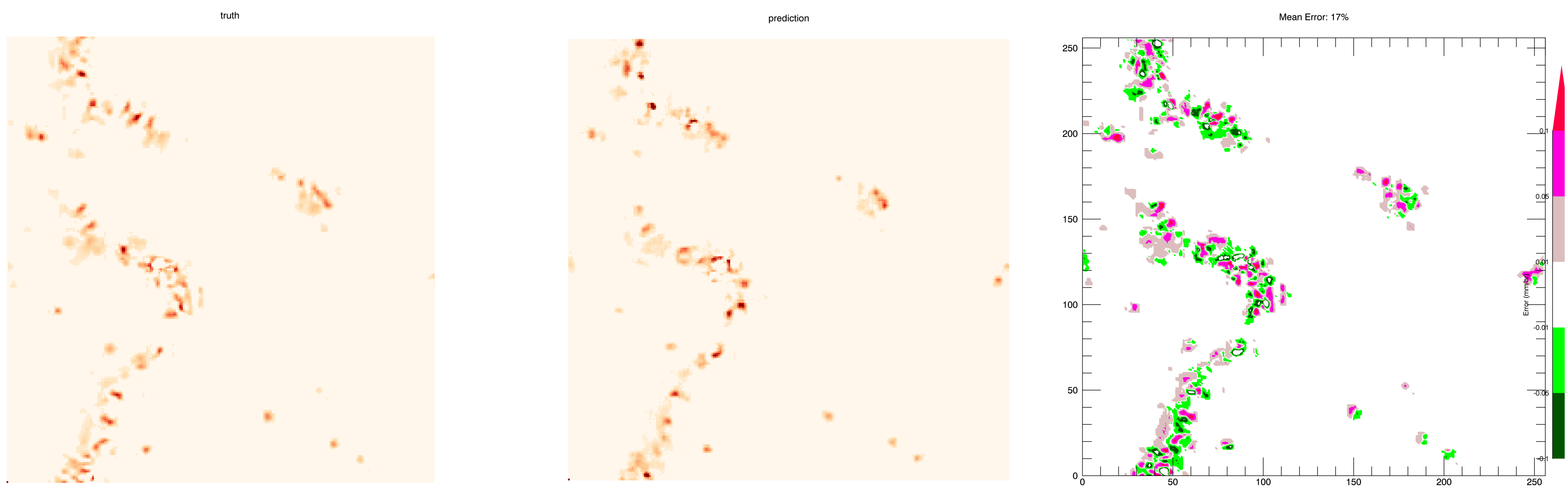
Simulation Performance

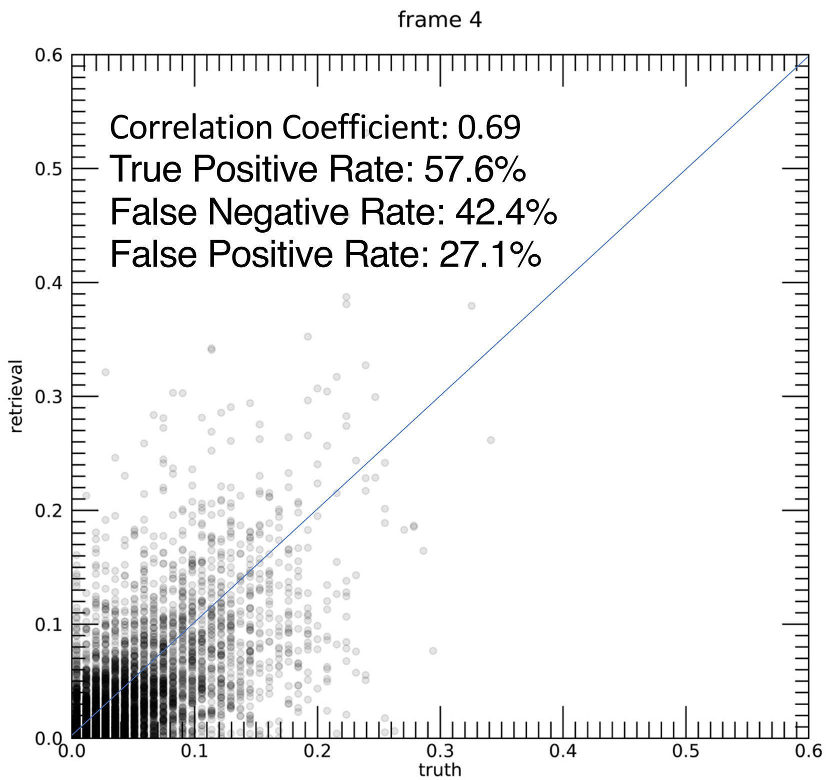
- One epoch takes ~10 minutes and 1000 epochs takes ~7 days
- Running 200 epochs produces the prediction images close to the truth images inspected with eyeballs which indicates the convergency is fast
- The results with 1000 epochs are similar to that with 2000 epochs. That means that this kind of simulations converges reasonably well.
  - The following analysis will use 5 outputs out of 200 outputs in a simulation run with 1000 epochs

One output from cGAN (Pix2pix) at 1000 epochs (Frame 4)



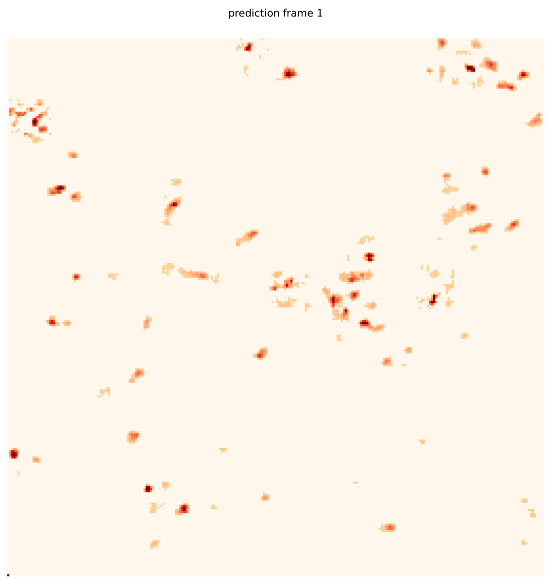
True, Prediction, Error for Organized Rainfall Systems

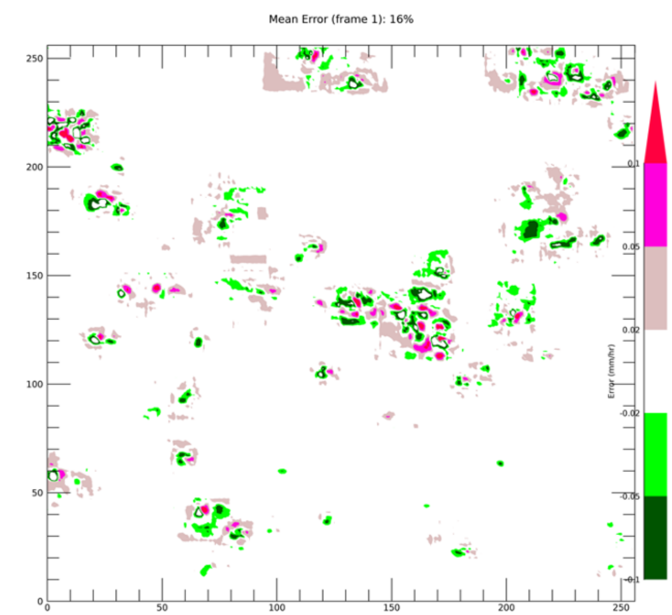


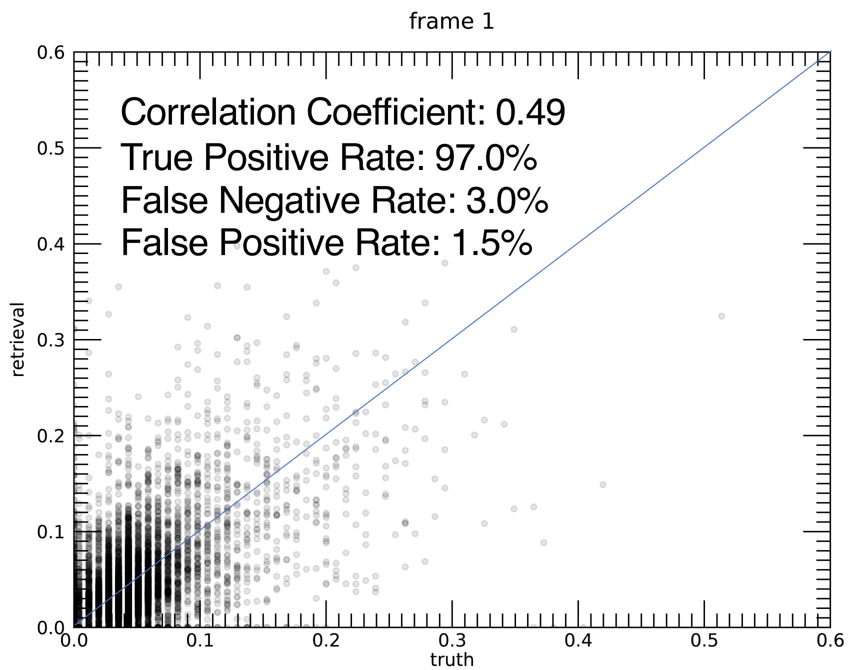


Note: The dots are in 90% transparent to see where most of the values are (each point has the same color value and darkness reflect the effect of overlapping points

True, Prediction, Error for Patchy Precipitation



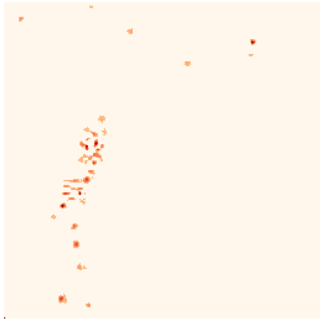
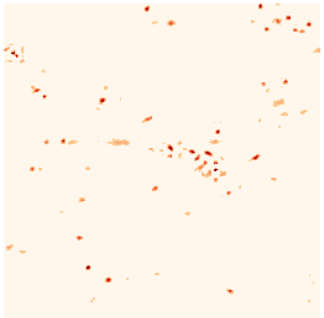
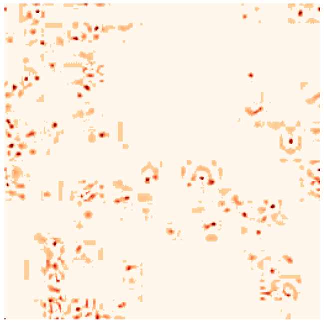
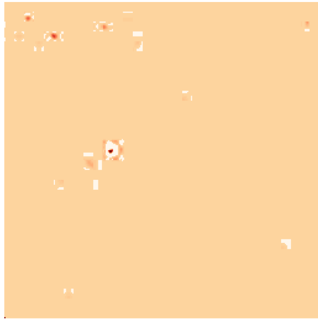
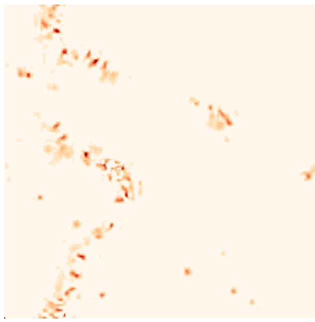




frame 1
Correlation Coefficient: 0.49
True Positive Rate: 97.0%
False Negative Rate: 3.0%
False Positive Rate: 1.5%
retrieval
truth
0.0
0.1
0.2
0.3
0.4
0.5
0.6

Performance Metrics

Frame	Correlation coefficient	True Positive	False Negative	False Positive	Total pixel
0	0.25	99.9%	0.1%	0.2%	65404
1	0.49	97.0%	3.0%	1.5%	64097
2	0.40	99.0%	1.0%	3.2%	63192
3	0.58	100.0%	0.0%	0.1%	65498
4	0.69	57.6%	42.4%	27.1%	3977
All	0.86	98.7%	1.3%	1.4%	262168

Overall Analysis

- For 5 (combined) outputs, the instantaneous surface rainfall rate retrieval has the correlation coefficient of 0.865. The true positive rate is 98.7%, whereas the false negative rate is 1.3%. The false positive percentage is 1.4%.
- These statistics are comparable to the National Mosaic and Multi-Sensor QPE (NMQ) system surface rain retrieval, and PERSIANN satellite rainfall retrieval, although cautions are needed because these products are not directly comparable to our study.**

**Zhang et al., 2011: National Mosaic and Mult-sensor QPE (NMZ) System: Description, Results, and Future Plans, *BAMS.*
Sadeghi et al., 2020: Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. *Environ. Modelling and Software.*

National Mosaic and Mult-sensor QPE (NMZ) System

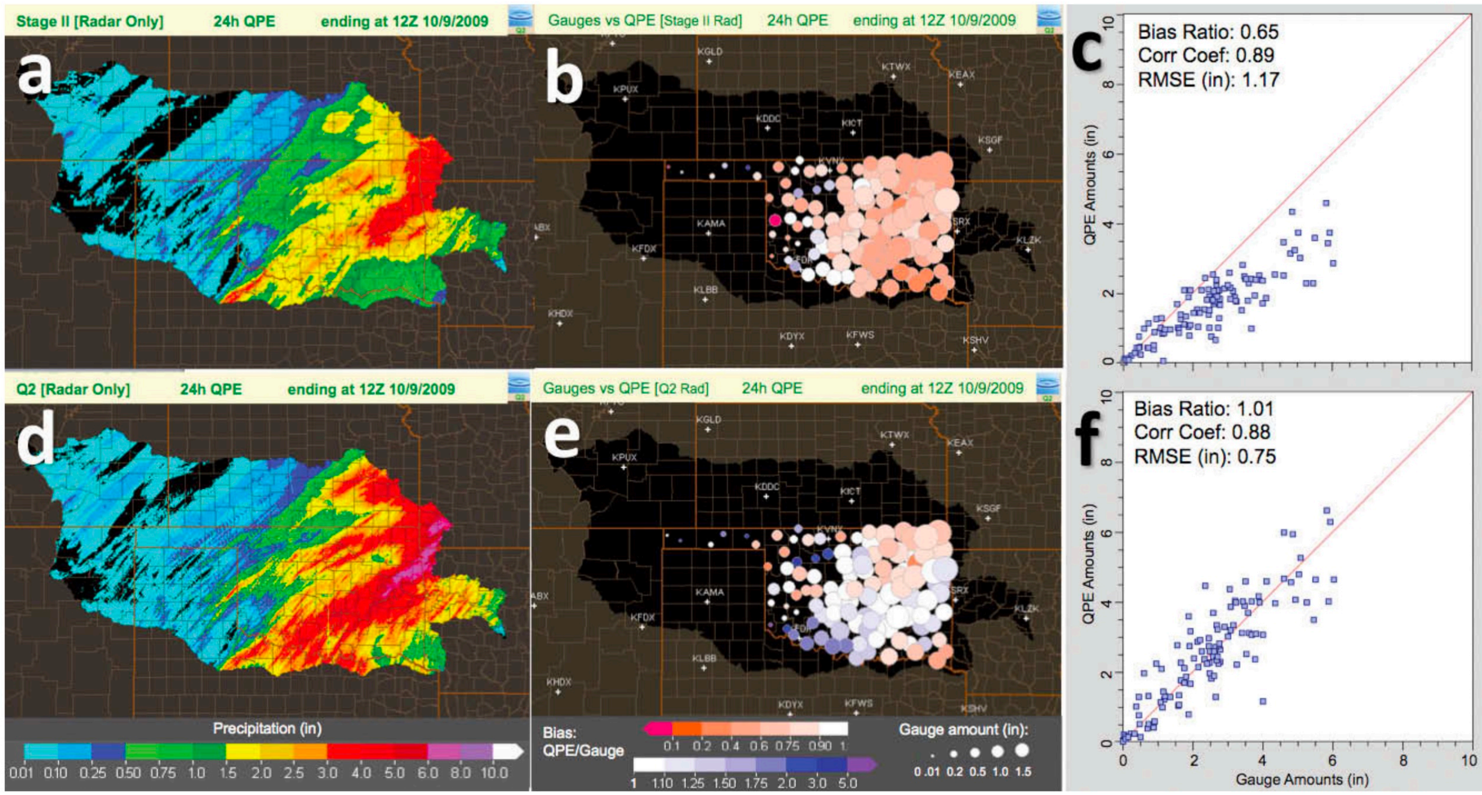


Fig. 11. Daily (a) stage II and (d) Q2 radar-only QPE maps ending at 1200 UTC 9 Oct 2009 in the ABRFC region. Bubble charts show bias ratios between (b) stage II and (e) Q2 radar-only QPEs and gauge observations, where the size of the circles represents the gauge-observed rainfall amount and the color shows the bias. Scatterplots show distributions of the 24-h (c) stage II and (f) Q2 radar-only QPEs vs the gauge observations.

PERSIANN satellite rainfall retrieval

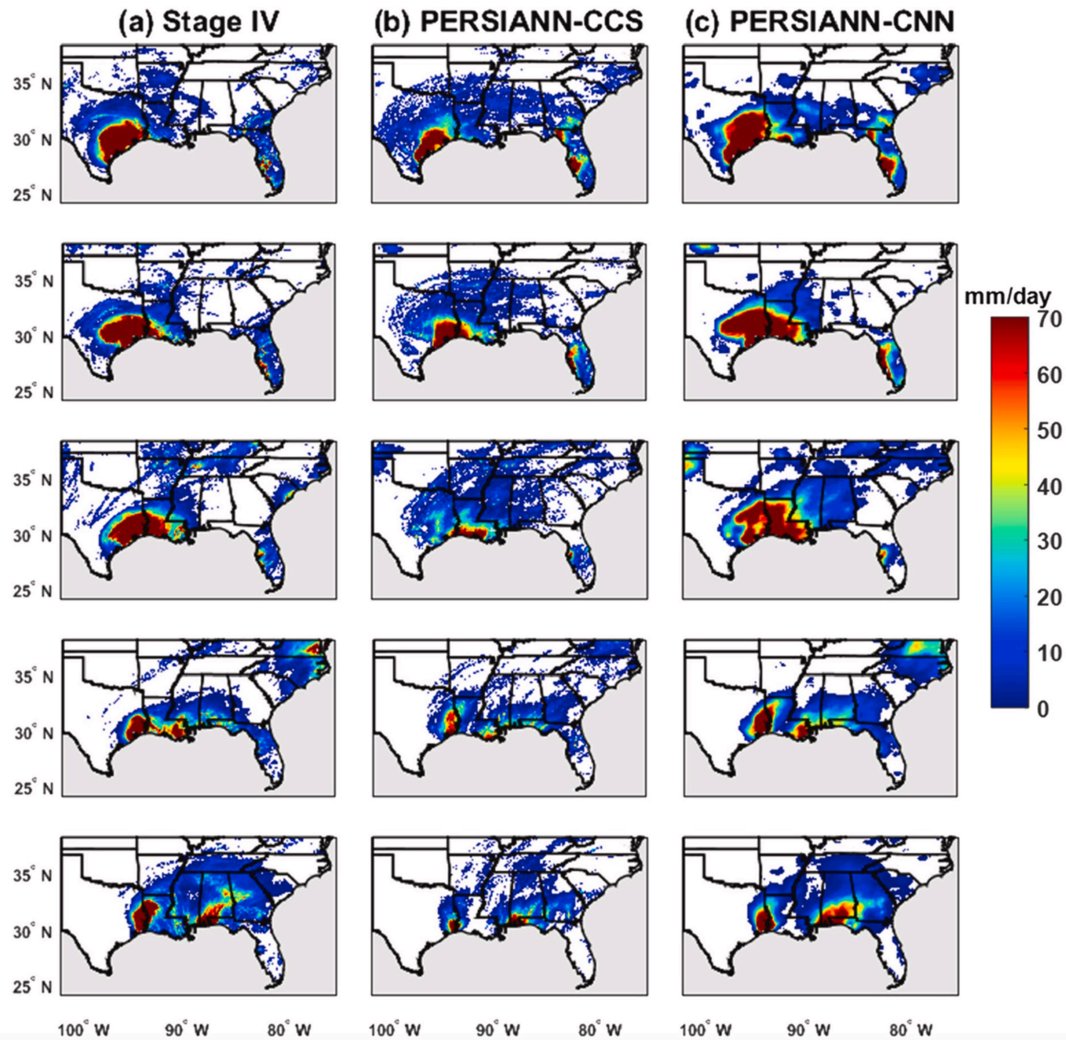


Fig. 7. The spatial precipitation evolution of Hurricane Harvey for the period between August 26 to 30 from (a) Stage IV, (b) PERSIANN-CCS, (c) PERSIANN-CNN

PERSIANN satellite rainfall retrieval

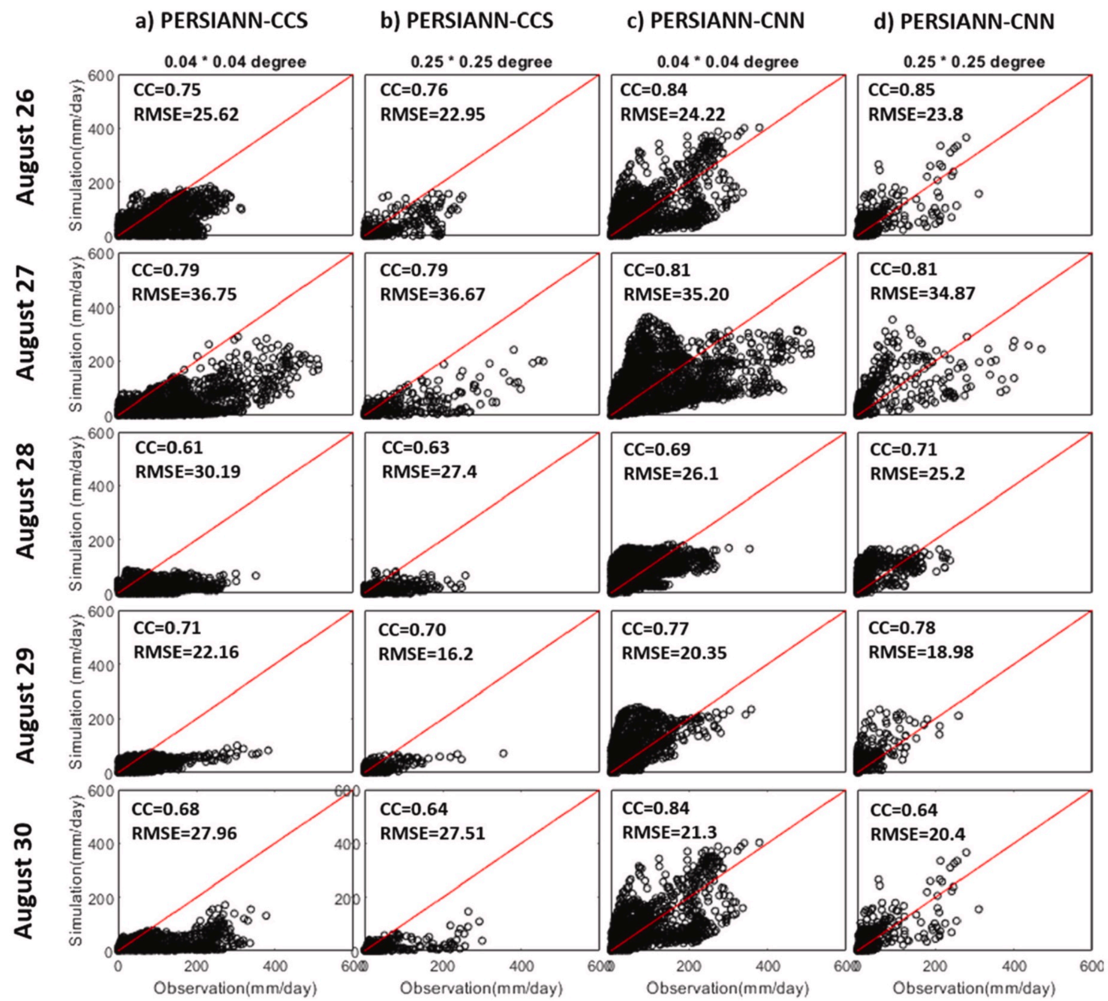


Fig. 8. Scatterplots of radar measurements (observation) vs PERSIANN-CCS, and PERSIANN-CNN (simulations): daily rainfall estimation at two spatial scales over the affected area during the Hurricane Harvey (Aug. 26 to 30, 2017).

Discussion

- The algorithm is better at identifying raining pixels when surface rainfall is scattered with low intensity. But the correlation coefficients are low due to large noises at the low rainfall rates.
- When there are organized rainfall systems (e.g., the mesoscale convective systems) and high rain intensity, the correlation coefficients improves, but the true positive rate becomes lower due to mis-identifications.
- In this study, 5 outputs are analyzed. Analysis of more outputs are needed

Conclusions

- The direct application of the cGAN algorithm shows promising overall results in mapping the radar images to rainfall images. However, the performance of the algorithm appears not uniform. More systematic simulations and analysis are needed.